The Strong Law of Large Numbers and the Glivenko-Cantelli Theorem in a finitely additive setting

by

Thomas E. Armstrong

University of Minnesota

Technical Report No. 446

November 1984

Introduction

Chen and Ramakrishnan [1983] give necessary and sufficient conditions in the finitely additive setting for a sequence of independent and identically distributed random variables to satisfy the Glivenko-Cantelli Theorem. Here it is shown that their conditions are equivalent to the random variables having the same distribution functions as a random variable on a countably additive measure space. It is shown that the validity of the Glivenko-Cantelli theorem holds for any independent sequence of random variables with respect to a product strategic measure (not necessarily a power measure) with the same distribution functions (not necessarily identically distributed) iff the common distribution functions are those of a random variable on a countably additive probability space. In doing so the Kolmogoroff strong law of large numbers must be extended from the setting of Chen [1981] to the present setting.

1) Distribution Functions

Let X_1 and X_2 be sets equipped with algebras $\mathcal{B}_1$ and $\mathcal{B}_2$ of subsets and finitely additive probabilities μ_1 and μ_2 respectively. Let $\mathcal{J}$ be the smallest algebra of subsets of $\mathbb{R} = (-\infty,\infty)$ containing the intervals $(-\infty,t)$ and $(-\infty, t]$ for all t. Let Z_1 and Z_2 be random variables on X_1 and X_2 in that $Z_i^{-1}(\mathcal{J}) \subset \mathcal{B}_i$ for $i=1,2$. Define the distribution functions $F_{Z_i}(t) = \mu_i(Z_i^{-1}(-\infty, t])$ and $G_{Z_i}(t) = \mu_i(Z_i^{-1}(-\infty, t))$ for $i=1,2$. If $F_{Z_1} = F_{Z_2}$ and $G_{Z_1} = G_{Z_2}$ say that Z_1 and Z_2 have the same distribution functions. If $(X_2,\mathcal{B}_2,\mu_2)$ is a countably additive probability space and Z_1 has the same distribution functions as Z_2 say that Z_1 has countably additive distribution.

Lemma 1.1 Z_1 and Z_2 have the same distribution functions iff the image measures $\mu_1 o Z_1^{-1}$ and $\mu_2 o Z_2^{-1}$ on $\mathcal{J}$ are the same.

Proof Immediate. □

Remark Call Z_1 and Z_2 identically distributed iff $\mathcal{B}_i$ contains $\sigma(Z_i^{-1}(\mathcal{J})) = \sigma(Z_i)$ for 1,2 and $\mu_i o Z_i^{-1}$ yields the same measure on $\sigma(\mathcal{J})$, the Borel sets in $(-\infty,\infty)$, for $i=1,2$, Karandikar [1982]. An even stronger definition is used by Chen and Ramakrishnan [1983] where $\mathcal{B}_i = 2^{X_i}$ for $i=1,2$ and they require $\mu_i o Z_i^{-1}$ to be the same measure on $2^{\mathbb{R}}$ for $i=1,2$.

Proposition 1.2 The following are equivalent if X is a set with $\mathcal{B}$ an algebra of subsets, μ a finitely additive probability on $\mathcal{B}$ and Z is a random variable on X.

i) Z has countably additive distribution.

ii) μ extends in a countably additive fashion from $Z^{-1}(\mathcal{J})$ to $\sigma(Z) = \sigma(Z^{-1}(\mathcal{J}))$.

iii) F_Z is right continuous, G_Z is left continuous and $o = F_Z(-\infty) = 1 - F_Z(\infty)$.

<u>Proof</u> i) $\Rightarrow$ iii) is immediate

iii) $\Rightarrow$ i) Let ν be the countably additive Borel measure on $(-\infty,\infty)$ with $F_Z(t) = \nu((-\infty, t])$ and $G_Z(t) = \nu(-\infty, t)$. Let $Z_o(x) = x$ be considered as a random variable on $(-\infty,\infty)$. Z_o and Z have the same distribution functions.

<u>iii) $\Rightarrow$ ii)</u> Take ν as in the proof of iii) $\Rightarrow$ i. Define $\hat{\mu}$ on $\sigma(Z^{-1}(J)) = Z^{-1}(\sigma(J))$ by $\hat{\mu}(Z^{-1}(A)) = \nu(A)$. $\hat{\mu}$ is a countably additive extension (by Lemma 1.1) of μ from $Z^{-1}(J)$ to $\sigma(Z^{-1}(J))$.

<u>ii) $\Rightarrow$ i)</u> F_Z and G_Z remain the distribution functions of Z if B is replaced by $\sigma(Z)$ and μ by the $\hat{\mu}$ of the proof of iii) $\Rightarrow$ ii). □

<u>Example 1.3</u> There exists a finitely additive Borel measure ν on $(-\infty,\infty)$ giving measure 1 to the dyadic rationals D in [0,1] so that $\nu([0,t]) = t$ for $0 \leq t \leq 1$.

To see this let δ_x denote Dirac measure on the Borel sets for real x. Let ν_n be $2^{-n} \sum_{k=1}^{2^n} \delta_{k2^{-n}}$. Consider $\{\nu_n : n \in N\}$ to be a net of finitely additive Borel measures. Let ν be any limit point of this net for the (compact) topology of pointwise convergence on the Borel sets. Since $\nu_n(D) = 1$ for all n we have $\nu(D) = 1$. Since $\nu_m([0,k2^{-n}]) = k2^{-n}$ for all $k \leq 2^n$ and $n \leq m$ we have $\nu([0,k2^{-n}]) = k2^{-n}$ if $k \leq 2^n$ for all n. By continuity we have $\nu([0,t]) = t$ for all $t \in [0,1]$.

<u>Remarks</u> It may be shown that if D is a countable dense set in $(-\infty,\infty)$ and ν_2 is a countably additive Borel probability measure on $(-\infty,\infty)$ there exists a finitely additive Borel probability measure ν_1 on $(-\infty,\infty)$ with $\nu_1(D) = 1$ and $\nu_1((-\infty, t]) = \nu_2((-\infty, t])$ for all t. Letting $Z_1(t) = Z_2(t)$ for all t one

obtains Z_1 and Z_2 with the same distribution functions. Z_2 has countably additive distribution even though ν_2 is purely finitely additive. Example 1.3 shows this in case ν is uniform distribution on [0,1].

This example shows that in Proposition 1.2 ii) may hold without μ being countably additive on $\sigma(Z)$ even if $\sigma(Z) \subset \mathcal{B}$.

We may extend ν from the Borel sets of $(-\infty,\infty)$ to $2^{\mathbb{R}}$ without affecting the conclusion of Example 1.3.

2) Kolmogoroff's Strong Law of Large Numbers

Let X be a discrete space so that 2^X is the relevant algebra of sets. Let $H = X^\infty$ be given the product topology. Let $\{\gamma_n : n \in N\}$ be a sequence of probability measures on X thought of as forming a strategy $\sigma = (\sigma_o, \sigma_1, \ldots \sigma_n, \ldots)$ with $\sigma_o = \gamma$ and $\sigma_{n-1}(x_1,\ldots,x_{n-1},dx_n) = \gamma_n(dx_n)$ for all $(x_1,\ldots,x_n) \varepsilon X^n$. Associate to σ the strategic measure σ on the clopen algebra of H as in Dubins and Savage [1965] and extended by regularity to a σ-algebra by Purves and Sudderth [1976]. The strategic measure σ thus obtained is called a strategic product measure and is denoted by $\gamma_1 \times \gamma_2 \times \ldots \times \gamma_n \times \ldots$. The strategy σ is termed an independent strategy in this case. If $\gamma_j = \gamma$ for all j we speak of a strategic power measure and denote it by γ^∞. A sequence of random variables $\{Z_n : n \in N\}$ on X gives rise to a corresponding sequence of coordinate functions $\{Z_n : n \varepsilon N\}$ on H by the requirement that $Z_k(x_1,\ldots,x_n,\ldots) = Z_k(x_k)$.

Theorem (Kolmogoroff's Strong Law of Large Numbers)

Let γ be the strategic product measure $\gamma_1 \times \ldots \times \gamma_n \times \ldots$. Let $\{Z_n : n \in N\}$

be a sequence of coordinate functions. Suppose that with respect to γ_n the distribution functions F_{Z_n} and G_{Z_n} are F and G for all n. $n^{-1}(Z_1 + \ldots + Z_n)$ converges γ-almost surely to the common mean $\mu = \int_X Z_n \, d\gamma_n$ iff Z_1 is γ_1 integrable.

<u>Proof</u> The integrability of Z_1 or any other Z_n depends only on the distribution functions F and G. Indeed the proof of Lemma 4.5 of Chen [1977] shows that Z_1 is integrable iff $\sum_{n=1}^{\infty}[1 - F(n) + G(-n) + G(-n)] < \infty$. When Z_1 is integrable then $\int Z_1 d\gamma_1$ is expressible in terms of F and G. To see this for $-\infty < t < \infty$ define H(t) to be $F(t^+)$. Note that $F(-\infty) = H(-\infty) = 0 = 1 - H(\infty) = 1 - F(\infty)$ since Z_1 is integrable. One may easily verify that $\int Z_1 d\gamma_1 = \int_{-\infty}^{\infty} t \, dH(t)$.

Assume that Z_1 is integrable and set $Y_n = Z_n - \mu$ for all n. It must be shown that $n^{-1}(Y_1 + \ldots + Y_n)$ converges to 0 γ-almost surely. To establish this we follow classical reasoning as in Rao [1984]. We first note that Chen [1977] Corollary 4.5 shows that this holds provided that $\sum_{n=1}^{\infty} n^{-2}\left[\int_X Y_n{}^2 \, d\gamma_n\right] < \infty$, even if $\{Y_n\}$ fail to have common distribution functions. Even if $\{Z_n\}$ fail to have common distribution functions but

$\mu_n = \int_X Z_n \, d\gamma_n \to \mu$ then $n^{-1}(Z_1 + \ldots + Z_n) \to \mu$ γ-almost surely as $n \to \infty$ as

long as $\sum_{n=1}^{\infty} n^{-2}\left[\int_X Z_n^2 \, d\gamma_n - \int_X Z_n \, d\gamma n^{\,2}\right] < \infty$, Chen [1977] Corollary 4.1

Let $E_n = \{|X_n| > n\}$, $F_n = \{|Y_n| \leq n\}$, $U_n = Y_n \chi_{F_n}$ and $V_n = Y_n \chi_{E_n}$ so $Y_n = U_n + V_n$ for all n. Note that $\mu_n = \int_X U_n d\gamma_n = \int_{\{|Y_1| \geq n\}} Y_1 d\gamma_1$ converges to $\int_X Y_1\, d\gamma_1 = 0$ as $n \to \infty$. Next note that $\int_X U_n^2\, d\gamma_n - \mu_n^2 \leq \int_{\{|X_1| \leq n\}} X_1^2\, d\gamma_1$ $\leq \sum_{k=0}^{n-1} (k+1)^2\, \gamma_1(\{k < |Y_1| \leq k+1\})$. Thus,

$$\sum_{n=1}^{\infty} n^{-2}\left[\int_X U_n^{\,2}\, d\gamma_n - \mu^2_{\,n}\right] \leq \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} \frac{(k+1)^2}{n^2}\, \gamma_1(\{k < |Y_1| \leq k+1\})$$

$$= \sum_{k=1}^{\infty} k^2\, \gamma_1(\{k-1 < |Y_1| \leq k\}) \sum_{n=k}^{\infty} \frac{1}{n^2}$$

$$\leq \sum_{k=1}^{\infty} k^2\, \gamma_1(\{k-1 < |Y_1| \leq k\})\; 2/k$$

$$= 2 \sum_{k=1}^{\infty} k\gamma_1(\{k-1 < |Y_1| \leq k\}) \leq 2\int_X |Y_1|\, d\gamma_1 + 2 < \infty .$$

Thus, as previously noted, $\frac{1}{n}(U_1 + \ldots + U_n) \to 0$ γ-almost as surely as $n \to \infty$. To show that $\frac{1}{n}(Y_1 + \ldots + Y_n) \to 0$ γ-almost surely it is now sufficient to show that $\frac{1}{n}(V_1 + \ldots + V_n) \to 0$ γ - almost surely. Actually it is the case for γ-almost all $x = (x_n) \varepsilon H$ that $V_n(x_n) = 0$ for n sufficiently large. This is a consequence of the Borel-Cantelli Corollary 2.2 of Chen [1977] and the fact that $\sum_{n=1}^{\infty} \gamma_n(\{V_n \neq 0\}) = \sum_{n=1}^{\infty} \gamma_1\{|Y_1| > n\} < \infty$. The latter sum is finite by Lemma 4.5 of Chen [1977] since Y_1 is γ_1-integrable.

One implication of the theorem has been established. For the converse assume that $(Z_1 + \ldots + Z_n)n^{-1} = S_n\, n^{-1}$ converges γ-almost everywhere to μ. Write $Z_n\, n^{-1}$ as $S_n\, n^{-1} - S_{n-1}(n-1)^{-1}(n-1)n^{-1}$ to deduce that $Z_n\, n^{-1} \to 0$ γ-almost surely. The Borel-Cantelli Corollary 2.2 of Chen [1977] shows that $\sum_{n=1}^{\infty} \gamma_n(\{|Z_n| > n\}) = \sum_{n=1}^{\infty} \gamma_1(\{|Z_1| > n\}) < \infty$. This shows that $\int_X |Z_1|\, d\gamma_1 < \infty$

by Lemma 4.5 of Chen [1977]. This completes the proof of the theorem. □

Corollary 2-2-1 For each $n \in N$ let X_n be a set, $\mathcal{B}_n$ an algebra of subsets, γ_n a finitely additive probability measure on X_n and Z_n a random variable on X_n with $F_{Z_n} = F$ and $G_{Z_n} = G$. Extend each γ_n arbitrarily to 2^{X_n} and let $\gamma = \gamma_1 \times \ldots \times \gamma_n \ldots$ be the strategic product measure (which may be constructed by considering γ_n defined on 2^X where $X = X_1 \cup \ldots \cup X_n \ldots$) In order that the strong law of large numbers hold for $\{Z_n\}$ it is necessary and sufficient that Z_1 be γ_1 integrable.

Remark The event $E = \{(x): \frac{Z_1(x_1)+\ldots+Z_n(x_n)}{n} \to \int_X Z_1 d\gamma_1\}$ is in the tail σ-field determined by the random variables $\{Z_n : n \in N\}$. This corollary merely states that $\gamma(E) = 1$ for all strategic product measures whose n^{th} factor agrees with γ_n on $\mathcal{B}_n$. The most important case in Corollary 2-2-1 is when $\mathcal{B}_n = Z_n^{-1}(\mathcal{J})$ for all n. Next most important is the case $\mathcal{B}_n = \sigma(Z_n)$.

3) The Glivenko-Cantelli Theorem

Let $\{\gamma_n\}$ be a sequence of finitely additive probability measures on 2^X and let $\{Z_n\}$ be an associated sequence of random variables on X with common distribution functions $F = F_{Z_n}$ and $G = G_{Z_n}$. Let γ be the strategic product measure $\gamma_1 \times \gamma_2 \times \ldots \times \gamma_n \times \ldots$. For each n construct the empirical distribution functions $G_n(t,x) = \sum_{j=1}^{n} n^{-1} \chi_{(-\infty,\, t)}(Z_j(x_j))$ and

$F_n(t,x) = \sum_{j=1}^{n} n^{-1} \chi_{(-\infty,\, t]}(Z_j(x_j))$. Set $Y_n^-(t,x) = \chi_{(-\infty,\, t)}(Z_n(x))$ and

$Y_n^+(t,x) = \chi_{(-\infty,\, t]}(Z_n(x))$. For any n $F_{Y_n^+(t_i,\cdot)}(s) = \begin{cases} 0 & s < 0 \\ 1 - F(t) & 0 \le s < 1 \\ 1 & 1 \le s \end{cases}$ and

$$G_{Y_n^+(t,\cdot)}(s) = \begin{cases} 0 & s \leq 0 \\ 1 - F(t) & 0 < s \leq 1 \\ 1 & 1 \leq s \end{cases}$$

. Thus, $\left\{Y_n^+(t,\cdot)\right\}$ have common distribution functions with $\int_X Y_1^+(t,\cdot)d\gamma_1 = \gamma_1(\{Z_1 \leq t\}) = F(t)$. Similarly $\left\{Y_n^-(t,\cdot)\right\}$ have common distribution functions with $\int_X Y_1^-(t,\cdot)d\gamma_1 = \gamma_1(\{Z_1 < t\}) = G(t)$. Application of Proposition 2-2 to the sequences $\left\{Y_n^-(t,\cdot)\right\}$ and $\left\{Y_n^+(t,\cdot)\right\}$ of integrable random variables immediately yields this proposition. □

<u>Proposition 3.1</u> For any $-\infty < t < \infty$

$$\lim_{n\to\infty} G_n(t,x) = G(t) \text{ and } \lim_{n\to\infty} F_n(t,x) = F(t)$$

for γ-almost all x in H.

The Glivenko-Cantelli Theorem is concerned with the set C of $x \in H$ so that the empirical distribution functions $F_n(\cdot,x)$ converge uniformly to $F(\cdot)$ on $(-\infty,\infty)$ rather than merely pointwise. Since each $F_n(\cdot,x)$ is a distribution function of a countably additive probability it follows that so is F and that $F(t^-) = \lim_{n\to\infty} F_n(t-,x)$ if $t \in \mathbb{R}$ and $x \in C$.

<u>Proposition 3.2</u> If $\gamma(C) > 0$ then the random variables $\{Z_j\}$ have common countably additive distribution.

<u>Proof</u> It must be shown that $F(t) - F(t^-) = \gamma_j(\{t\})$ for each point of discontinuity t of F. Select one. For γ-almost all x, hence γ-almost all x in C, hence for at least one $x \in C$ the strong law of large numbers tells us that $F_n(t,x) - F_n(t^-,x) = F_n(t,x) - G_n(t,x) = \sum_{j=1}^{n} \frac{1}{n} \chi_{\{t\}}(Z_j(x_j))$ converges

to $\gamma_1(\{t\}) = F(t) - G(t)$. But also $F_n(t,x) - F_n(t^-, x)$ converges to $F(t) - F(t^-)$. Since t is arbitrary the proposition is established. □

Proposition 3.3 If $\{Z_j\}$ has a common countably additive distribution then $\gamma(C) = 1$.

Proof Pick $\varepsilon > 0$. Pick $t_m < t_M$ so that $F(t_m) < \varepsilon/2$ and $F(t_M) \geq 1 - \varepsilon/2$. For γ-almost all $x \varepsilon H$ there is an $n(x)$ so that $|F_n(t_m,x) - F(t_m)| \leq \varepsilon/2$ and $|F_n(t_M,x) - F(t_M)| \leq \varepsilon/2$ for $n \geq n(h)$. Thus $|F_n(t,x) - F(t)| \leq \varepsilon$ if $t \varepsilon [t_m, t_M]$ and $n \geq n(h)$.

Let D denote a countable dense set in $[t_m, t_M]$ containing all discontinuity points of F. For each $d \varepsilon D$ there is a tail event E_d of γ-probability 1 in H so that if $x \varepsilon E_d$ then $F_n(d,x)$ converges to $F(d)$ and if d is a discontinuity point of F then $F_n(d,x) - F_n(d^-,x)$ approaches $\gamma_1(\{d\}) = F(d) - F(d^-)$. Lemma 1 of Chen and Ramakrishnan assures us that if $x \varepsilon E = \cap \{E_d : d \varepsilon D\}$ then $F_n(\cdot,x)$ converges uniformly to $F(\cdot)$ on $[t_m, t_M]$. Since γ is countably additive on the tail algebra by Purves and Sudderth [1977] or [1983] it follows that $\gamma(D) = 1$. Thus, for γ-almost all $x \varepsilon H$ $F_n(\cdot,\alpha)$ converges uniformly to $F(\cdot)$ on $[t_m, t_M]$. Since the set of x in H which have $F_n(\cdot,x)$ eventually within ε of $F(\cdot)$ outside $[t_m, t_M]$ is also a tail event of γ-probability 1 it follows that for γ-almost all $x \varepsilon H$ $F_n(\cdot,x)$ converges uniformly to $F(\cdot)$ outside $[t_m, t_M]$. This suffices to establish the proposition. □

Putting Propositions 3.2 and 3.3 together we have the Glivenko-Cantelli Theorem.

Glivenko-Cantelli Theorem Let X_j be a set, $\mathcal{B}_j$ an algebra of sets γ_j a finitely additive probability measure on $\mathcal{B}_j$ and Z_j a random variable on X_j with distribution functions $F_{Z_j} = F$ and $G_{Z_j} = G$ for all $j \varepsilon N$. Extend γ_j to 2^X where $X = {}_j\cup X_j$ and let γ be the associated strategic product measure. In order that $F_n(\cdot, z)$ converge uniformly to $F(\cdot)$ it is necessary and sufficient that the Z_j's have countably additive distribution. In this case $G_n(\cdot, x)$ converges uniformly to $G(\cdot)$ for γ-almost all $x \varepsilon H$.

Remark One may interchange the roles of F and G in the Glivenko-Cantelli Theorem.

References

Chen, R. [1977]. On almost sure convergence in a finitely additive setting, Zeit. Warsch. Verw. Geb, 37, 341-356.

Chen, R. [1981]. A finitely additive version of the Glivenko-Cantelli theorem, Bull. Inst. Math. Academica Sinica, 9, 303-310.

Chen, R., Ramakrishnan, S. [1983]. On the Glivenko-Cantelli Theorem in a finitely additive setting, Sankhya, The Indian J. Stat, 45, 161-167.

Dubins, L., Savage, L.J.[1965]. Inequalities for Stochastic Processes. How to Gamble if You Must. McGraw Hill, New York.

Karandikar, R.L., [1982]. A general principle for limit theorems in finitely additive probability. Trans. Amer. Math. Soc. 273, 541-550.

Purves, R., Sudderth, W.D. [1976]. Some finitely additive probability, Ann. Prob.

Purves, R., Sudderth, W.D. [1983]. Finitely additive 0-1-Laws, Sankhya, The Indian J. Statistics, 45, 32-37.

Rao, M.M. [1984]. Probability Theory with Applications, Academic Press, New York.